

INICIAR SESIÓN

NUESTROS PLANES

TODOS LOS
CURSOS

FORMACIONES

CURSOS

PARA
EMPRESAS

ARTÍCULOS DE TECNOLOGÍA

Manipulación de strings en pandas: lower, replace, startswith y contains



Guilherme Silveira

18/05/2022



Esse artigo faz parte da
Formação Data Science



Imagina una plataforma de música en línea llamada Alurafy, que te permite reproducir canciones, listas de reproducción y suscribirte a dos tipos de planes. Los usuarios acceden a las páginas (URIs) y tenemos los datos en bruto en una tabla con: • columna que

representa la ID de usuario (si está conectado) • columna con el URI accedido • columna con el estado de respuesta, si fue un éxito (200) o un error interno (500):

```
datos = pd.read_csv("todos_accesos.csv")
datos.head()
```

	URI	Usuario logado	Status de respuesta
0	/	16629.0	200
1	/musicas	614.0	200
2	/contato	9420.0	200
3	/playlist/drumnbass	7566.0	200
4	/	6070.0	200

Básicamente, cada dataframe tiene un conjunto de datos que son strings y, a menudo, no nos damos cuenta, cuáles son los nombres de las columnas: URI, Usuario conectado y Status de respuesta:

```
datos.columns
```

Resultando en:

```
Index(['URI', 'Usuario conectado', 'Estado de respuesta'], dtype='object')
```

Son nombres razonablemente descriptivos, pero horribles de acceder. Por ejemplo, no pudimos acceder a:

```
datos.Usuario conectado
```

Pero solamente:

```
datos['Usuario conectado']
```

Por supuesto, no es una gran pérdida. Pero tenemos mayúsculas y minúsculas mezcladas, no hay un estándar claro. ¿Qué tal estandarizar todas estas strings? Para esto tomamos los nombres de las columnas y mediante el atributo `str` logramos ponerlos todos en letra minúsculas:

```
datos.columns.str.lower()
```

Resultando en columnas en minúsculas:

```
Index(['uri', 'usuario conectado', 'estado de respuesta'], dtype='object')
```

Pero los espacios en blanco siguen ahí. Podemos tomar esta serie de datos, usar de nuevo el str y reemplazar los espacios con guiones bajos:

```
datos.columns.str.lower().str.replace(' ', '_')
```

Por fin, las columnas con los nombres que quería:

```
Index(['uri', 'usuario_conectado', 'estado_de_respuesta'], dtype='object')
```

Si asignamos los valores de estas columnas a sí mismas:

```
datos.columns = datos.columns.str.lower().str.replace(' ', '_')
datos.head()
```

	uri	usuario_logado	status_de_resposta
0	/	16629.0	200
1	/musicas	614.0	200
2	/contato	9420.0	200
3	/playlist/drumnbass	7566.0	200
4	/	6070.0	200

Y esta línea de estandarización de nombres de columnas se puede usar en casi todos los proyectos inmediatamente después de importar un csv.

Ahora echemos un vistazo a las páginas accesadas a través de tus URI:

```
datos.uri.unique()
```

```
array(['/', '/canciones', '/contacto', '/playlist/drumnbass',
      '/playlist/funk', '/compra/plan_basico', '/playlist/techno',
      '/promocion_primer_de_abril/compra/plan_basico',
      '/compra/plan_avanzado', '/playlist/pop', '/playlist/rock',
      '/playlist/jazz', '/playlist/clasica', '/playlist/kpop'],
      dtype=object)
```

Ahora me gustaría crear una nueva columna indicando si al acceder a esta página el usuario quería comprar algo. Es decir, ¿la URL comienza con /compra/? Si es, True, si no False. Para esto tenemos la función `startswith`:

```
datos['comprando'] = datos.uri.str.startswith("/compra")
```

Así que estamos buscando en una columna de string/texto las líneas que comienzan con un determinado valor. Ahora imprimimos con una query los primeros 5 elementos que están con True:

```
datos.query("comprando==True").head()
```

Resultando en:

	uri	usuario_logado	status_de_resposta	comprando
13	/compra/plano_basico	7031.0	200	True
31	/compra/plano_avancado	2147.0	200	True
37	/compra/plano_avancado	7948.0	200	True
86	/compra/plano_avancado	11177.0	200	True
104	/compra/plano_avancado	8775.0	200	True

Para agrupar y contar el número de páginas de compra versus el número de otras páginas, tomemos la columna `comprando` y sumemos sus valores:

```
datos.comprando.value_counts()
```

Y notamos que tenemos alrededor del 9,5% de visitas a las páginas de compras:

```
False 164295
```

```
True 15705
```

```
Name: comprando, dtype: int64
```

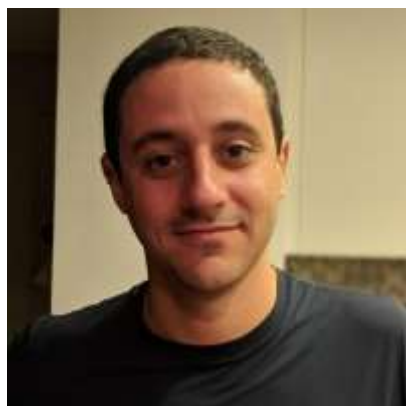
Por fin, cometimos un error. De hecho, también hay páginas de compra como del día 1 de abril en `/promocion_primeiro_de_abril/compra/plano_basico`. Por lo tanto, no solo queremos las URI que comienzan con `/compra/` pero que tienen en el medio el tramo `/compra`. Además de la función `startswith`, `str` nos permite realizar varias otras funciones, como `contains`:

```
datos['comprando'] = datos.uri.str.contains("/compra/")
datos.query("comprando==True").head()
```

	uri	usuario_logado	status_de_resposta	comprando
13	/compra/plano_basico	7031.0	200	True
24	/promocao_primeiro_de_abril/compra/plano_basico	14887.0	200	True
31	/compra/plano_avancado	2147.0	200	True
33	/promocao_primeiro_de_abril/compra/plano_basico	4871.0	200	True
37	/compra/plano_avancado	7948.0	200	True

Ahora sí, creamos la nueva columna de acuerdo con cualquier página que tenga el tramo `/compra/` en su URI.

En resumen, siempre que necesites trabajar con una columna de tipo string, para extraer valores de ella, hacer transformaciones de string a string, echa un vistazo a la documentación de la `str` a ver si ya tiene lo que quieres hacer. Hay decenas de funciones.



Guilherme cofundó Caelum, Alura y GUJ. Con más de 15 años de experiencia enseñando programación y habilidades digitales, coordina los equipos de producción de cursos en

Alura. Es un tecnólogo con sesgo matemático y medallista de oro en competencias nacionales de computación, representando a Brasil en los mundiales. Participante activo en la comunidad open source y de educación en tecnología, autor de 7 libros en el área.



Esse artigo faz parte da
Formação Data Science

ARTÍCULOS DE TECNOLOGÍA

En Alura encontrarás variados cursos sobre . ¡Comienza ahora!

SEMESTRAL

US\$49,90

un solo pago de US\$49,90

- ✓ 218 cursos
- ✓ Videos y actividades 100% en Español
- ✓ Certificado de participación
- ✓ Estudia las 24 horas, los 7 días de la semana
- ✓ Foro y comunidad exclusiva para resolver tus dudas
- ✓ Acceso a todo el contenido de la plataforma por 6 meses

¡QUIERO EMPEZAR A ESTUDIAR!

[Paga en moneda local en los siguientes países](#)

ANUAL

US\$79,90

un solo pago de US\$79,90

- ✓ 218 cursos
- ✓ Videos y actividades 100% en Español
- ✓ Certificado de participación
- ✓ Estudia las 24 horas, los 7 días de la semana
- ✓ Foro y comunidad exclusiva para resolver tus dudas
- ✓ Acceso a todo el contenido de la plataforma por 12 meses

¡QUIERO EMPEZAR A ESTUDIAR!

[Paga en moneda local en los siguientes países](#)

Acceso a todos
los cursos

Estudia las 24 horas,
dónde y cuándo quieras

Nuevos cursos
cada semana

NAVEGACIÓN

PLANES

INSTRUCTORES

BLOG

POLÍTICA DE PRIVACIDAD

TÉRMINOS DE USO

SOBRE NOSOTROS

PREGUNTAS FRECUENTES

¡CONTÁCTANOS!

¡QUIERO ENTRAR EN CONTACTO!

BLOG

PROGRAMACIÓN

FRONT END

DATA SCIENCE

INNOVACIÓN Y GESTIÓN

DEVOPS

AOVS Sistemas de Informática S.A

CNPJ 05.555.382/0001-33

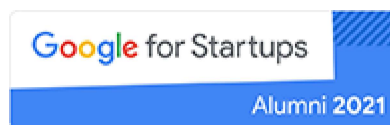
SÍGUENOS EN NUESTRAS REDES SOCIALES



ALIADOS



En Alura somos unas de las Scale-Ups seleccionadas por Endeavor, programa de aceleración de las empresas que más crecen en el país.



Fuimos unas de las 7 startups seleccionadas por Google For Startups en participar del programa Growth Academy en 2021

CURSOS

Cursos de Programación

Lógica de Programación | Java

Cursos de Front End

HTML y CSS | JavaScript | React

Cursos de Data Science

Data Science | Machine Learning | Excel | Base de Datos | Data Visualization | Estadística

Cursos de DevOps

Docker | Linux

Cursos de Innovación y Gestión

Productividad y Calidad de Vida | Transformación Ágil | Marketing Analytics |

Liderazgo y Gestión de Equipos | Startups y Emprendimiento